ISSN 1990-4789, Journal of Applied and Industrial Mathematics, 2010, Vol. 4, No. 1, pp. 79–93. © Pleiades Publishing, Ltd., 2010.
Original Russian Text © B.Yu. Lemeshko, S.B. Lemeshko, S.N. Postovalov, 2008, published in Sibirskii Zhurnal Industrial'noi Matematiki, 2008, Vol. XI, No. 4, pp. 78–93.

Comparative Analysis of the Power of Goodness-of-Fit Tests for Near Competing Hypotheses. II. Verification of Complex Hypotheses

B. Yu. Lemeshko*, **S. B. Lemeshko****, and **S. N. Postovalov*****

Novosibirsk State Technical University, pr. K. Marksa 20, Novosibirsk, 630092 Russia

Received November 25, 2007; in final form, February 21, 2008

Abstract—In this part of the article, we use the statistical modeling methods to analyze the power of a series of goodness-of-fit tests for complex hypotheses. We estimate the power of tests relative to some near competing hypotheses. The combination of results enables us to order the tests with respect to their power for testing both simple and complex hypotheses.

DOI: 10.1134/S1990478910010114

Key words: *goodness-of-fit test, Kolmogorov test, Cramér–von Mises–Smirnov omega-squared test, Anderson–Darling test, Pearson test, Nikulin test, power of a test*

INTRODUCTION

In this article which completes [1], we present the results of studying the power of a series of goodness-of-fit tests in the case of testing complex hypotheses. On the same pairs of competing hypotheses we analyze the power of the Kolmogorov, Cramér–von Mises–Smirnov, Anderson–Darling, Pearson χ^2, and Nikulin χ^2-type tests.

For testing a complex hypothesis H_0 of the form $F(x) \in \{F(x,\theta), \theta \in \Theta\}$, where Θ is the domain of the unknown parameter θ, we consider the situations when the parameter $\hat{\theta}$ of the theoretical distribution is estimated using the same sample on which goodness-of-fit is tested.

Recall that, in these situations, the nonparametric Kolmogorov, Cramér–von Mises–Smirnov, and Anderson–Darling tests lose their "freeness from distributions." The distributions $G(S|H_0)$ of the statistics of the tests depend on the form of the observed distribution $F(x,\theta)$ corresponding to the hypothesis H_0 being tested; on the type of the estimated parameter and the number of parameters estimated using the sample; in some situations, on the concrete value of the parameter(s) (for instance, in the case of gamma- and beta-distributions); and on the method used for estimating the parameters. Also, the distribution of the statistics of the Pearson χ^2 test in the case of estimating θ by using ungrouped data is not a χ^2-distribution.

For testing complex hypotheses with all goodness-of-fit tests under study, in order to estimate the unknown parameters we use the maximal likelihood method. In this case, on the one hand, all tests are on equal footing; and, on the other hand, the nonparametric tests of Kolmogorov, Cramér–von Mises–Smirnov ω^2, and Anderson–Darling Ω^2 type when we use maximal likelihood estimates (MLE) have higher power in comparison with the case when we find estimates in result of minimizing the statistics of the corresponding test [2, 3].

As in [1], we consider two pairs of competing hypotheses. In the first pair, the tested hypothesis H_0 corresponds to the normal law with density

$$f(x) = \frac{1}{\theta_0\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta_1)^2}{2\theta_0^2}\right\},$$

*E-mail: lemeshko@fpm.ami.nstu.ru
**E-mail: skyer@mail.ru
***E-mail: Postovalov@ngs.ru

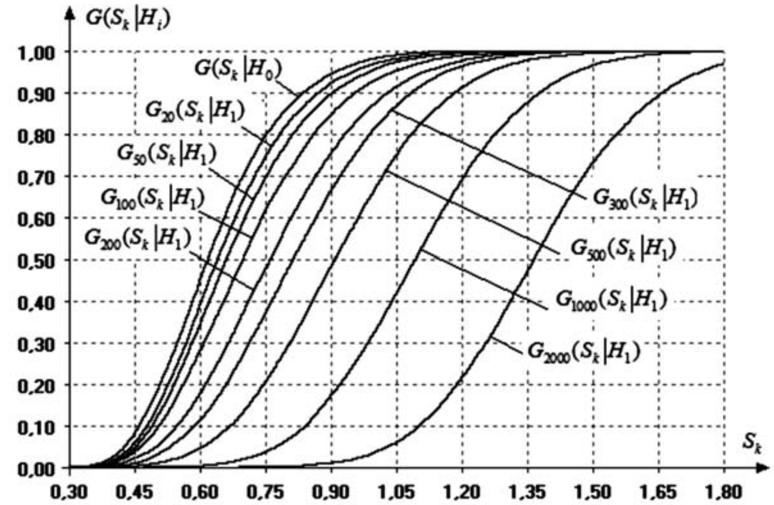

Fig. 1. The distributions $G(S_k|H_0)$ and $G_n(S_k|H_1)$ of the statistics of the Kolmogorov test for testing complex hypotheses H_0 on goodness-of-fit with the normal law in the case of using MLE for the competing hypothesis H_1 corresponding to a logistic law

and the competing hypothesis H_1, to the logistic density

$$f(x) = \frac{\pi}{\theta_0\sqrt{3}} \exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\} \Big/ \left[1 + \exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\}\right]^2,$$

while $\theta_0 = 1$ and $\theta_1 = 0$.

In the second pair, H_0 corresponds to the Weibull distribution with density

$$f(x) = \frac{\theta_0(x-\theta_2)^{\theta_0-1}}{\theta_1^{\theta_0}} \exp\left\{-\left(\frac{x-\theta_2}{\theta_1}\right)^{\theta_0}\right\}$$

and the translation parameter $\theta_2 = 0$, and H_1, to the gamma-distribution with density

$$f(x) = \frac{1}{\theta_1\Gamma(\theta_0)} \left(\frac{x-\theta_2}{\theta_1}\right)^{\theta_0-1} e^{-(x-\theta_2)/\theta_1},$$

while $\theta_0 = 3.12154$, $\theta_1 = 0.557706$, and $\theta_2 = 0$.

We study the power for testing the complex hypotheses H_0 against the simple hypotheses H_1.

1. THE POWER OF TESTS IN THE CASE OF COMPLEX HYPOTHESES AND THE ALTERNATIVE "NORMAL DISTRIBUTION vs. LOGISTIC"

Figure 1 depicts the results of modeling the distribution of the Kolmogorov statistics with the Bol'shev correction for testing the complex hypothesis on goodness-of-fit with a normal law with estimating the parameters of the normal law using the maximal likelihood method for the competing hypothesis H_1 corresponding to a logistic law. The figure depicts the distribution $G(S_k|H_0)$ and the functions $G_n(S_k|H_1)$ for the sample volumes of $n = 20$, 50, 100, 300, 500, 1000, and 2000 observations. As we see, for the same confidence level $\alpha = 0.1$, the power of the test turns out considerably higher than for testing the simple hypothesis [1]; and it is about 0.142 for $n = 20$, 0.181 for $n = 50$, 0.236 for $n = 100$, 0.351 for $n = 200$, 0.459 for $n = 300$, 0.646 for $n = 500$, 0.905 for $n = 1000$, and 0.997 for $n = 2000$. This indicates that, in the case of testing complex hypotheses, the same laws may differ for average sample volumes.

Figure 2 depicts the distributions $G(S_\omega|H_0)$ and $G_n(S_\omega|H_1)$ of the Cramér–von Mises–Smirnov S_ω statistics for testing the complex hypothesis and the same H_0 and H_1 with calculating the MLE for the parameters of the normal law. Fig. 3 shows the analogous distributions $G(S_\Omega|H_i)$ for the Anderson–

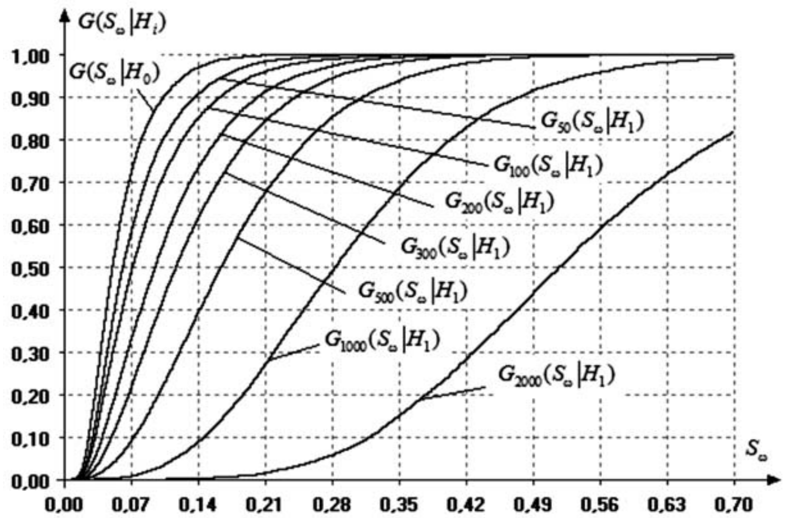

Fig. 2. The distributions of the statistics $G(S_\omega|H_0)$ of the Cramér–von Mises–Smirnov ω^2 test and $G_n(S_\omega|H_1)$ for testing the complex hypothesis H_0 on goodness-of-fit with a normal law in the case of using MLE for the competing hypothesis H_1 corresponding to a logistic law

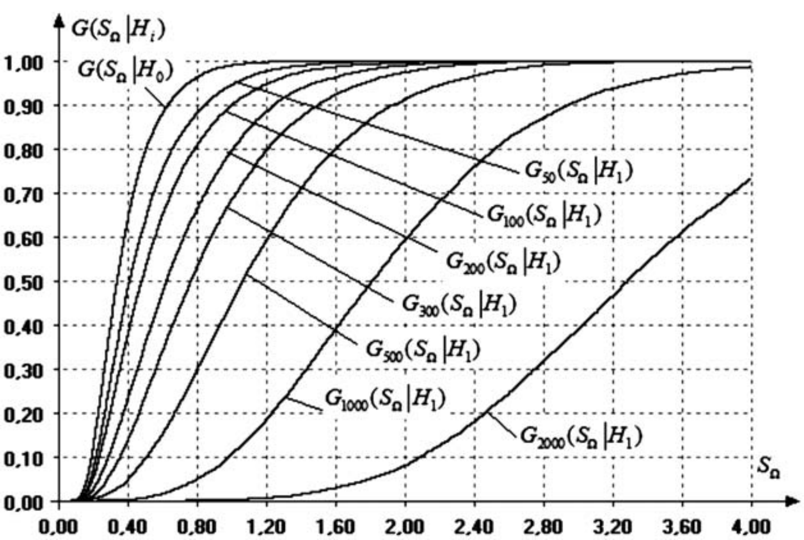

Fig. 3. The distributions of the statistics $G(S_\Omega|H_0)$ of the Anderson–Darling ω^2 test and $G_n(S_\Omega|H_1)$ for testing the complex hypothesis H_0 on goodness-of-fit with a normal law in the case of using MLE for the competing hypothesis H_1 corresponding to a logistic law

Darling S_Ω statistics.

Figure 4 depicts the distributions of Y_n^2 statistics of the Nikulin test for testing the complex hypothesis and calculating the MLE for two parameters of the normal law using the ungrouped data. The picture corresponds to the case of using the asymptotically optimal grouping (AOG) with $k = 9$ intervals.

We should note that the distributions $G(Y_n^2|H_0)$ of the Nikulin statistics Y_n^2 converge quite fast to the corresponding χ^2_{k-1}-distributions. We observe a substantial difference from χ^2_{k-1}-distributions only for small sample volumes and large numbers of intervals. For small sample volumes, the distributions of the statistics show the effect of the finiteness of the number of intervals.

As an example, Table 1 presents the attained confidence levels for testing goodness-of-fit of the empirical distributions of Y_n^2 statistics with χ^2_{k-1}-distributions for $n = 100$ and $n = 500$, which indicate

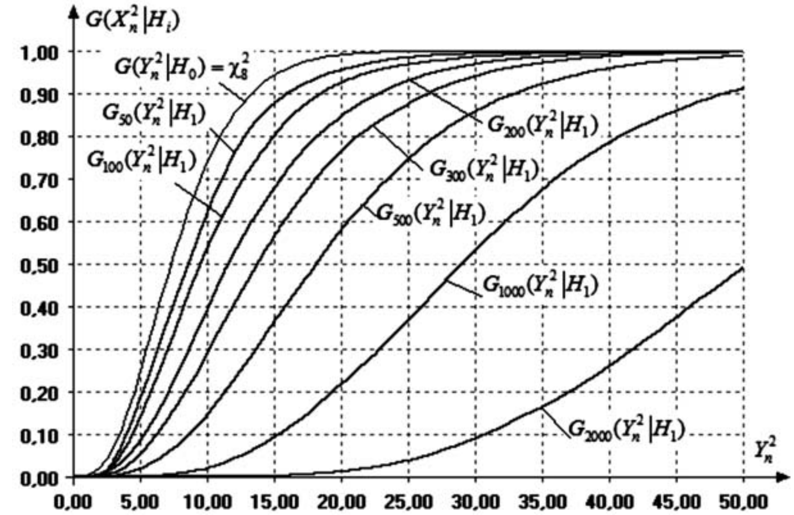

Fig. 4. The distributions of the Nikulin statistics Y_n^2, $G(Y_n^2|H_0) = \chi_8^2$ and $G(Y_n^2|H_1)$, for testing the complex hypothesis H_0 on goodness-of-fit with a normal law and with the competing hypothesis H_1 in the case of AOG with $k = 9$ and basing the MLE on the ungrouped data

some satisfactory degree of proximity of the distributions for $n = 100$ and a high degree for $n = 500$. The hypotheses were tested using the samples of volume $N = 10000$.

Table 2 presents the estimates calculated basing on the results of modeling the distributions of the statistics, for the power of goodness-of-fit tests under consideration and for various values of the confidence level α, for testing the complex hypothesis H_0 corresponding to the normal law against the hypothesis H_1 corresponding to the logistic law with the parameters $(0, 1)$. In this table the tests are also ordered in decreasing power.

Note that, in some cases, the preference is not obvious since, while possessing a greater power for one confidence level and one sample volume, a test may lose out for other values of α and n.

Table 2 presents the maximal power of Nikulin's and Pearson's χ^2 tests (for given grouping methods). Table 3 presents the value of the power of these tests for other numbers of intervals with AOG and equiprobable grouping (EPG).

We emphasize that, estimating the power for testing complex hypotheses, we use the modeled distributions $G(S|H_0)$ of the statistics for the sample volume $n = 1000$. For n so large, the empirical distribution of the statistics can be assumed to be a good estimate for the limit law. We did that in analyzing the power of nonparametric goodness-of-fit tests and Pearson's χ^2 test. To estimate the power of the test with the Nikulin statistics we used its available limit, the χ_{k-1}^2-distribution.

Table 1. The attained confidence levels for testing goodness-of-fit of the modeled distributions of Y_n^2 statistics with χ_{k-1}^2-distributions

Goodness-of-fit tests	$k = 9$		$k = 15$	
	$n = 100$	$n = 500$	$n = 100$	$n = 500$
Pearson χ^2	0.0180	0.9732	0.1905	0.2037
Kolmogorov	0.1817	0.8951	0.4450	0.9480
von Mises ω^2	0.1334	0.9755	0.2943	0.8425
Anderson–Darling ω^2	0.0948	0.9579	0.2974	0.7988

Table 2. The power of goodness-of-fit tests for the complex hypothesis H_0 (normal distribution) against the competing hypothesis H_1 (logistic)

α	$n = 20$	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
	the power of the Anderson–Darling Ω^2 test							
0.15	0.222	0.297	0.400	0.575	0.708	0.873	0.989	1.000
0.10	0.164	0.230	0.324	0.496	0.636	0.828	0.981	1.000
0.05	0.098	0.149	0.224	0.377	0.519	0.741	0.963	1.000
0.025	0.060	0.096	0.152	0.282	0.414	0.649	0.935	0.999
0.01	0.031	0.054	0.091	0.186	0.297	0.525	0.885	0.998
	the power of the Nikulin χ^2-type test with $k = 15$ and AOG							
0.15	0.245	0.320	0.395	0.536	0.646	0.806	0.967	1.000
0.10	0.195	0.249	0.332	0.466	0.579	0.755	0.952	0.999
0.05	0.137	0.165	0.248	0.368	0.480	0.669	0.921	0.998
0.025	0.077	0.112	0.184	0.291	0.395	0.587	0.883	0.996
0.01	0.036	0.071	0.125	0.213	0.304	0.488	0.825	0.992
	the power of the Cramér–von Mises–Smirnov ω^2-type test							
0.15	0.210	0.273	0.366	0.529	0.659	0.836	0.980	1.000
0.10	0.153	0.208	0.291	0.447	0.582	0.781	0.968	1.000
0.05	0.090	0.130	0.194	0.329	0.458	0.678	0.939	0.999
0.025	0.053	0.082	0.128	0.237	0.353	0.573	0.897	0.998
0.01	0.027	0.044	0.074	0.150	0.243	0.445	0.825	0.994
	the power of the Pearson χ^2 test with $k = 15$ and AOG							
0.15	0.243	0.295	0.342	0.467	0.579	0.751	0.950	0.999
0.10	0.194	0.220	0.280	0.393	0.502	0.688	0.928	0.998
0.05	0.140	0.133	0.199	0.291	0.391	0.583	0.882	0.996
0.025	0.081	0.080	0.137	0.214	0.303	0.486	0.827	0.992
0.01	0.036	0.043	0.079	0.139	0.213	0.376	0.745	0.984
	the power of the Kolmogorov test							
0.15	0.200	0.246	0.313	0.440	0.554	0.732	0.941	0.999
0.10	0.142	0.181	0.236	0.351	0.459	0.646	0.905	0.997
0.05	0.080	0.105	0.143	0.230	0.322	0.502	0.823	0.990
0.025	0.045	0.061	0.086	0.149	0.219	0.376	0.721	0.975
0.01	0.021	0.029	0.043	0.081	0.127	0.244	0.575	0.938

Table 3. The power of Nikulin's and Pearson's χ^2 goodness-of-fit tests for the complex hypothesis H_0 (normal distribution) against the competing hypothesis H_1 (logistic distribution) depending on the grouping method and number of intervals

α	$n=20$	$n=50$	$n=100$	$n=200$	$n=300$	$n=500$	$n=1000$	$n=2000$
	the power of Nikulin's χ^2 type test with $k=9$ and AOG							
0.15	0.203	0.269	0.339	0.480	0.599	0.774	0.958	0.999
0.10	0.144	0.204	0.271	0.406	0.525	0.714	0.940	0.999
0.05	0.080	0.129	0.186	0.304	0.417	0.617	0.901	0.997
0.025	0.046	0.084	0.129	0.227	0.329	0.525	0.854	0.994
0.01	0.023	0.049	0.081	0.155	0.239	0.419	0.784	0.988
	the power of Nikulin's χ^2 type test with $k=9$ and EPG							
0.15	0.200	0.258	0.333	0.477	0.600	0.782	0.964	1.0008
0.10	0.143	0.195	0.262	0.399	0.523	0.722	0.947	0.999
0.05	0.081	0.121	0.175	0.292	0.410	0.621	0.910	0.998
0.025	0.047	0.076	0.117	0.213	0.318	0.525	0.865	0.996
0.01	0.023	0.042	0.070	0.140	0.225	0.415	0.796	0.991
	the power of Nikulin's χ^2 type test with $k=15$ and EPG							
0.15	0.197	0.263	0.328	0.465	0.584	0.765	0.959	0.999
0.10	0.141	0.198	0.259	0.389	0.508	0.705	0.940	0.999
0.05	0.081	0.126	0.174	0.286	0.399	0.604	0.901	0.998
0.025	0.048	0.080	0.119	0.211	0.311	0.512	0.855	0.995
0.01	0.024	0.045	0.072	0.140	0.222	0.404	0.785	0.990
	the power of Nikulin's χ^2 type test with $k=7$ and EPG							
0.15	0.199	0.256	0.332	0.476	0.598	0.780	0.962	0.999
0.10	0.145	0.193	0.260	0.397	0.522	0.719	0.944	0.999
0.05	0.082	0.120	0.172	0.290	0.407	0.616	0.905	0.997
0.025	0.047	0.075	0.114	0.210	0.314	0.520	0.858	0.995
0.01	0.023	0.041	0.066	0.136	0.219	0.406	0.785	0.989
	the power of Nikulin's χ^2 type test with $k=7$ and AOG							
0.15	0.189	0.250	0.320	0.457	0.573	0.751	0.948	0.999
0.10	0.136	0.187	0.249	0.380	0.497	0.688	0.926	0.998
0.05	0.077	0.115	0.164	0.276	0.386	0.584	0.880	0.995
0.025	0.042	0.072	0.110	0.201	0.298	0.490	0.826	0.991
0.01	0.018	0.040	0.066	0.133	0.210	0.382	0.746	0.982

Table 3. (Contd.)

α	$n = 20$	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
	the power of Pearson's χ^2 test with $k = 9$ and AOG							
0.15	0.197	0.237	0.282	0.404	0.516	0.699	0.930	0.998
0.10	0.136	0.171	0.210	0.323	0.431	0.624	0.898	0.997
0.05	0.071	0.096	0.130	0.219	0.314	0.504	0.834	0.992
0.025	0.038	0.055	0.080	0.146	0.225	0.398	0.760	0.984
0.01	0.017	0.027	0.041	0.085	0.141	0.283	0.654	0.966
	the power of Pearson's χ^2 test with $k = 7$ and AOG							
0.15	0.178	0.213	0.265	0.382	0.490	0.672	0.913	1.000
0.10	0.124	0.154	0.195	0.299	0.403	0.589	0.874	0.999
0.05	0.067	0.087	0.113	0.195	0.284	0.462	0.797	0.998
0.025	0.038	0.048	0.066	0.126	0.197	0.356	0.711	0.995
0.01	0.016	0.022	0.032	0.069	0.118	0.242	0.593	0.989
	the power of Pearson's χ^2 test with $k = 7$ and EPG							
0.15	0.169	0.213	0.243	0.335	0.427	0.599	0.873	0.994
0.10	0.129	0.145	0.179	0.251	0.336	0.500	0.813	0.989
0.05	0.073	0.078	0.097	0.153	0.212	0.356	0.696	0.971
0.025	0.032	0.042	0.056	0.091	0.135	0.248	0.577	0.942
0.01	0.014	0.020	0.025	0.045	0.071	0.146	0.426	0.881
	the power of Pearson's χ^2 test with $k = 9$ and EPG							
0.15	0.171	0.198	0.227	0.309	0.390	0.541	0.828	0.989
0.10	0.127	0.142	0.167	0.232	0.296	0.445	0.757	0.979
0.05	0.068	0.070	0.092	0.136	0.187	0.309	0.629	0.951
0.025	0.037	0.038	0.049	0.079	0.187	0.208	0.502	0.908
0.01	0.013	0.018	0.022	0.039	0.060	0.120	0.358	0.828
	the power of Pearson's χ^2 test with $k = 15$ and EPG							
0.15	0.144	0.179	0.212	0.267	0.331	0.459	0.736	0.967
0.10	0.097	0.134	0.151	0.197	0.252	0.367	0.653	0.944
0.05	0.064	0.072	0.077	0.116	0.153	0.245	0.517	0.893
0.025	0.028	0.038	0.044	0.065	0.091	0.161	0.396	0.825
0.01	0.012	0.016	0.019	0.031	0.046	0.089	0.268	0.717

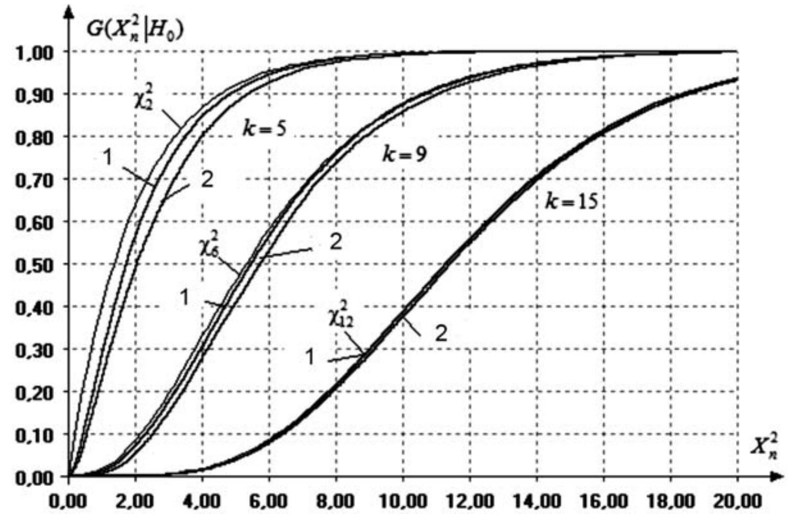

Fig. 5. The distributions of Pearson's X_n^2 statistics to χ^2_{k-r-1}-distributions for testing complex hypotheses and basing MLE on ungrouped data, depending on the grouping method AOG (curves 1) or EPG (curves 2)

Table 4. Model distributions of the statistics of the tests for testing complex hypotheses on goodness-of-fit with a normal distribution and the simultaneous estimation of two parameters by the maximal likelihood method

Goodness-of-fit tests	Model distribution of the statistics of tests
Anderson–Darling Ω^2	$G(S_\Omega\|H_0) = Su(-2.7377, 1.7186, 0.1023, 0.0980)$
Cramér–von Mises–Smirnov ω^2	$G(S_\omega\|H_0) = Sl(-2.5084, 1.8289, 0.1953, 0.0018)$
Kolmogorov	$G(S_K\|H_0) = \gamma(6.9372, 0.0560, 0.2486)$
Pearson's χ^2 with $k = 9$ and AOG	$G(X_n^2\|H_0) = \gamma(3.1121, 1.9551, 0.0422)$
Pearson's χ^2 with $k = 9$ and EPG	$G(X_n^2\|H_0) = \gamma(3.2465, 1.9313, 0.0980)$
Pearson's χ^2 with $k = 15$ and AOG	$G(X_n^2\|H_0) = \gamma(5.3451, 2.1221, 0.7022)$
Pearson's χ^2 with $k = 15$ and EPG	$G(X_n^2\|H_0) = \gamma(5.7292, 2.9454, 0.5000)$

The distributions $G(X_n^2|H_0)$ of X_n^2 statistics of the Pearson test for testing complex hypotheses and basing MLE on pointwise (ungrouped) data depend on the grouping method and differ from χ^2_{k-r-1}-distributions. For small numbers of intervals, these differences are considerable; and, as the number of intervals grows, the difference becomes of no practical value. Fig. 5 presents the distributions $G(X_n^2|H_0)$ of the statistics for $k = 5, 9, 15$ intervals while using AOG and EPG for estimating two parameters of the normal law and the corresponding χ^2_{k-r-1}-distributions. In the case of AOG, the distribution of the statistics $G(X_n^2|H_0)$ is always closer to the corresponding χ^2_{k-r-1}-distribution. For instance, for $k = 15$ and AOG, neither of the applied tests rejects anymore the hypothesis on the goodness-of-fit with the χ^2_{12}-distribution of the empirical distribution of the statistics corresponding to the modeled sample of volume $N = 20000$. At the same time, in the case of EPG, the similar hypotheses is rejected.

Table 4 presents the model distributions which approximate the limit distributions $G(S|H_0)$ of the statistics of the tests for testing the complex hypothesis H_0 relative to a normal distribution, and Table 5 presents percentage points.

In the tables we omit the Nikulin test whose limit distributions are χ^2_{k-1}-distributions. In Table 4, we

Table 5. Percentage points of the distributions of the statistics of the tests for testing complex hypotheses on goodness-of-fit with a normal distribution and the simultaneous estimation of two parameters by the maximal likelihood method

Goodness-of-fit tests	α				
	0.15	0.1	0.05	0.025	0.01
Anderson–Darling Ω^2	0.559	0.630	0.751	0.872	1.035
Cramér–von Mises–Smirnov ω^2	0.090	0.103	0.126	0.148	0.178
Kolmogorov	0.789	0.836	0.909	0.976	1.056
Pearson χ^2 with $k = 9$ and AOG	9.567	10.760	12.698	14.559	16.936
Pearson χ^2 with $k = 9$ and EPG	9.836	11.042	12.986	14.858	17.216
Pearson χ^2 with $k = 15$ and AOG	16.989	18.549	21.035	23.390	26.369
Pearson χ^2 with $k = 15$ and EPG	17.210	18.770	21.230	23.540	26.420

denote by $\gamma(\theta_0, \theta_1, \theta_2)$ the gamma-distribution with density

$$f(x) = \frac{1}{\theta_1^{\theta_0}\Gamma(\theta_0)}(x-\theta_2)^{\theta_0-1}e^{-(x-\theta_2)/\theta_1};$$

by $Sl(\theta_0, \theta_1, \theta_2, \theta_3)$, the Johnson distribution Sl with density

$$f(x) = \frac{\theta_1}{\sqrt{2\pi}\,(x-\theta_3)}\exp\left\{-\frac{1}{2}\left[\theta_0 + \theta_1 \ln\frac{x-\theta_3}{\theta_2}\right]^2\right\};$$

by $Su(\theta_0, \theta_1, \theta_2, \theta_3)$, the Johnson distribution Su with density

$$f(x) = \frac{\theta_1}{\sqrt{2\pi}\sqrt{(x-\theta_3)^2+\theta_2^2}}\exp\left\{-\frac{1}{2}\left[\theta_0 + \theta_1 \ln\left\{\frac{x-\theta_3}{\theta_2} + \sqrt{\left(\frac{x-\theta_3}{\theta_2}\right)^2+1}\right\}\right]^2\right\}.$$

We may indirectly judge the precision of modeling the empirical distributions of the statistics, for instance, by the essential coincidence of the resulting percentage points for the Cramér–von Mises–Smirnov ω^2 statistics with the corresponding values of these distributions tabulated for the normal law in [4].

In the case of testing complex hypotheses and sample volumes $n = 20$ and $n = 50$ for all tests under study, $G(S_{20}|H_0)$ and $G(S_{50}|H_0)$ differ considerably from $G(S_n|H_0)$ for $n = 1000$ and from χ^2_{k-1}-distributions for the Nikulin statistics. Thus, we estimated power using the modeled pairs of distributions of the form $G(S_{20}|H_0)$, $G(S_{20}|H_1)$, and $G(S_{50}|H_0)$, $G(S_{50}|H_1)$.

The Nikulin test is used for testing complex hypotheses. For it with EPG, for this pair of hypotheses, the power turns out maximal for an optimal number of intervals k that depends on the volume n of the sample [5]; and similarly for AOG. However, in the last case the optimal number k of intervals is shifted into the region of large values. In the case of this pair of competing hypotheses for large values of k, the use of AOG turns out preferable and, for smaller k (see $k = 7$), EPG wins out.

We can compare the power of goodness-of-fit tests for small sample volumes n to the power of tests constructed specifically for testing the deviation of distributions from the normal law: with the power of Shapiro–Wilk test, Epps–Pulley test, and D'Agostino test with the statistic z_2. Table 6 presents some estimates for the power of these tests for normality obtained in [6] and refined here for volumes of the modeled samples with $N = 10^6$. As we can see, the specific tests for the pair of hypotheses under consideration turn out more powerful.

Table 6. The power of tests for the deviation of distributions from the normal law (Shapiro–Wilk test, Epps–Pulley test, and D'Agostino test with the statistic z_2) relative to the competing hypotheses H_1 (logistic law)

α	normal law					
	Shapiro–Wilk		Epps–Pulley		D'Agostino z_2	
	$n = 20$	$n = 50$	$n = 20$	$n = 50$	$n = 20$	$n = 50$
0.1	0.181	0.202	0.178	0.249	0.189	0.327
0.05	0.117	0.141	0.111	0.165	0.111	0.223
0.01	0.044	0.067	0.037	0.062	0.032	0.089

2. THE POWER OF TESTS IN THE CASE OF COMPLEX HYPOTHESES FOR THE PAIR WEIBULL DISTRIBUTION vs. GAMMA-DISTRIBUTION

Tables 7 and 8 present the estimated power of the tests for various confidence levels α for testing the complex hypothesis H_0 on the goodness-of-fit of empirical distributions with a Weibull distribution, when using the sample we found MLE for the parameters θ_0 and θ_1 of this law for $\theta_2 = 0$, against the competing hypothesis H_1 corresponding to the gamma-distribution with the parameters $\theta_0 = 3.12154$, $\theta_1 = 0.557706$, and $\theta_2 = 0$. We order the tests in decreasing power.

Table 9 presents the model distributions approximating the limit distributions $G(S|H_0)$ of the statistics of the tests for testing the complex hypothesis H_0 relative to the Weibull distribution, and Table 10 presents percentage points. In the tables we omit the Nikulin test for which the limit distribution is known. In Table 9, we denote by $\ln N(\theta_1, \theta_0)$ the logarithmic normal distribution with density

$$f(x) = \frac{1}{x\theta_0\sqrt{2\pi}} e^{-(\ln x - \theta_1)^2/(2\theta_0^2)}.$$

Basing partially on the results of our previous studies [8], too harsh a conclusion is drawn in [7], which amounts to the recommendations for applying AOG in the χ^2-type tests being hardly valid. In this regard, we should note that, in both [8] and [9], we emphasize the unconditional positive effect of AOG in the case of near competing hypotheses for the Pearson χ^2 tests and the likelihood ratio, but we have not said that about the Nikulin test. Moreover, the results and conclusions there reveal a certain advantage of EPG. The results of these studies emphasize the more complicated nature of the dependence of the power of the Nikulin test (the Rao–Robson–Nikulin test) on the number of intervals and grouping methods and, in some cases, demonstrate a positive effect of AOG.

3. CONCLUSION

Combining the conclusions of [1] with these results, we can draw the following conclusions. For the case of testing simple hypotheses, we can order the tests under consideration with respect to their power as follows:

$$\text{Pearson (AOG)}\ \chi^2 \succ \text{ Anderson–Darling } \Omega^2 \succ \text{ von Mises } \omega^2 \succ \text{ Kolmogorov.}$$

We have this scale while using AOG in the Pearson χ^2 test, which minimizes the losses in the Fisher information. In the case of near competing hypotheses, the advantage in the power of the Pearson χ^2 test can be substantial.

For testing complex hypotheses, the order of preference turns out substantially different:

$$\text{Anderson–Darling } \Omega^2 \succ \text{ von Mises } \omega^2 \succ Y_n^2\ \text{(AOG)} \succ \text{ Pearson (AOG)}\ \chi^2 \succ \text{ Kolmogorov.}$$

For quite near hypotheses, it can be

$$\text{Anderson–Darling } \Omega^2 \succ Y_n^2\ \text{(AOG)} \succ \text{ von Mises } \omega^2 \succ \text{ Pearson (AOG)}\ \chi^2 \succ \text{ Kolmogorov.}$$

Table 7. The power of goodness-of-fit tests for the complex hypothesis H_0 (the Weibull distribution 2, 2, 0) with the competing hypothesis H_1 (the gamma-distribution with the parameters $3.12154, 0.557706, 0$)

α	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
	the power of the Anderson–Darling Ω^2 test					
0.15	0.435	0.667	0.817	0.952	0.999	1.000
0.10	0.353	0.589	0.757	0.928	0.998	1.000
0.05	0.244	0.466	0.650	0.876	0.995	1.000
0.025	0.167	0.361	0.547	0.811	0.990	1.000
0.01	0.100	0.252	0.424	0.715	0.977	1.000
	the power of the Cramér–von Mises–Smirnov ω^2 test					
0.15	0.396	0.603	0.750	0.913	0.996	1.000
0.10	0.316	0.520	0.679	0.875	0.993	1.000
0.05	0.212	0.394	0.560	0.797	0.984	1.000
0.025	0.143	0.295	0.452	0.712	0.968	1.000
0.01	0.082	0.196	0.330	0.593	0.936	1.000
	the power of the Nikulin χ^2 test with $k = 9$ and AOG					
0.15	0.324	0.511	0.665	0.869	0.993	1.000
0.10	0.246	0.423	0.584	0.818	0.987	1.000
0.05	0.153	0.299	0.454	0.720	0.973	1.000
0.025	0.096	0.209	0.347	0.619	0.951	1.000
0.01	0.051	0.129	0.238	0.492	0.909	0.999
	the power of the Pearson χ^2 test with $k = 9$ and AOG					
0.15	0.347	0.525	0.678	0.868	0.992	1.000
0.10	0.273	0.439	0.596	0.818	0.986	1.000
0.05	0.172	0.311	0.463	0.719	0.970	1.000
0.025	0.104	0.218	0.352	0.617	0.946	1.000
0.01	0.053	0.133	0.237	0.483	0.898	0.999
	the power of the Kolmogorov test					
0.15	0.340	0.510	0.646	0.830	0.981	1.000
0.10	0.262	0.420	0.558	0.762	0.965	1.000
0.05	0.164	0.293	0.420	0.640	0.925	0.999
0.025	0.101	0.200	0.306	0.519	0.867	0.997
0.01	0.052	0.115	0.193	0.375	0.763	0.988

Table 8. The power of Nikulin's and Pearson's χ^2 goodness-of-fit tests for the complex hypothesis H_0 (the Weibull distribution with 2, 2, 0) with the competing hypotheses H_1 (the gamma-distribution with 3.12154, 0.557706, 0) depending on the grouping method and the number of intervals

α	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
	the power of Nikulin's χ^2 test with $k = 15$ and AOG					
0.15	0.365	0.487	0.634	0.828	0.986	1.000
0.10	0.291	0.406	0.558	0.770	0.976	1.000
0.05	0.195	0.302	0.443	0.666	0.952	1.000
0.025	0.131	0.230	0.348	0.569	0.919	1.000
0.01	0.078	0.165	0.250	0.455	0.862	0.999
	the power of Pearson's χ^2 test with $k = 15$ and AOG					
0.15	0.385	0.491	0.637	0.839	0.988	1.000
0.10	0.305	0.405	0.559	0.780	0.980	1.000
0.05	0.196	0.293	0.442	0.676	0.959	1.000
0.025	0.124	0.218	0.344	0.574	0.928	1.000
0.01	0.065	0.151	0.234	0.448	0.868	0.999
	the power of Nikulin's χ^2 test with $k = 9$ and EPG					
0.15	0.295	0.455	0.599	0.806	0.981	1.000
0.10	0.220	0.367	0.509	0.740	0.968	1.000
0.05	0.133	0.250	0.378	0.624	0.938	1.000
0.025	0.080	0.167	0.276	0.512	0.894	0.999
0.01	0.040	0.097	0.176	0.380	0.822	0.997
	the power of Nikulin's χ^2 test with $k = 15$ and EPG					
0.15	0.273	0.421	0.558	0.774	0.975	1.000
0.10	0.202	0.335	0.468	0.702	0.960	1.000
0.05	0.120	0.224	0.341	0.582	0.923	1.000
0.025	0.071	0.147	0.244	0.469	0.874	0.999
0.01	0.036	0.085	0.153	0.343	0.796	0.997
	the power of Pearson's χ^2 test with $k = 9$ and EPG					
0.15	0.259	0.361	0.465	0.639	0.896	0.996
0.10	0.187	0.282	0.376	0.552	0.851	0.992
0.05	0.114	0.181	0.257	0.421	0.764	0.982
0.025	0.062	0.113	0.170	0.310	0.664	0.964
0.01	0.028	0.057	0.096	0.198	0.529	0.926

Table 8. (Contd.)

α	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
	the power of Pearson's χ^2 test with $k = 15$ and EPG					
0.15	0.252	0.348	0.454	0.638	0.910	0.998
0.10	0.179	0.268	0.364	0.553	0.869	0.996
0.05	0.099	0.165	0.244	0.415	0.784	0.990
0.025	0.053	0.102	0.158	0.303	0.688	0.978
0.01	0.026	0.052	0.089	0.196	0.559	0.955

Table 9. Model distributions of the statistics of the tests for testing complex hypotheses on the goodness-of-fit with a Weibull distribution and the simultaneous estimation of the shape and scale parameters by the maximal likelihood method

Goodness-of-fit tests	Model distributions of the statistic tests
Anderson–Darling Ω^2	$G(S_\Omega \mid H_0) = Sl(1.4612, 2.0543, 0.6704, 0.0165)$
Cramér–von Mises–Smirnov ω^2	$G(S_\omega \mid H_0) = LnN(-2.9747, 0.5320)$
Kolmogorov	$G(S_K \mid H_0) = \gamma(7.2210, 0.0535, 0.245)$
Pearson χ^2 with $k = 9$ and AOG	$G(X_n^2 \mid H_0) = \gamma(3.1809, 1.9248, 0)$
Pearson χ^2 with $k = 9$ and EPG	$G(X_n^2 \mid H_0) = \gamma(3.3897, 1.8906, 0)$
Pearson χ^2 with $k = 15$ and AOG	$G(X_n^2 \mid H_0) = \gamma(5.5355, 2.0817, 0.51)$
Pearson χ^2 with $k = 15$ and EPG	$G(X_n^2 \mid H_0) = \gamma(5.3077, 2.1504, 0.86)$

Table 10. Percentage points of the distributions of the statistics of the tests for testing complex hypotheses on the goodness-of-fit with a Weibull distribution and the simultaneous estimation of the shape and scale parameters by the maximal likelihood method

Goodness-of-fit tests	α				
	0.15	0.1	0.05	0.025	0.01
Anderson–Darling Ω^2	0.563	0.635	0.757	0.879	1.036
Cramér–von Mises–Smirnov ω^2	0.089	0.102	0.124	0.145	0.174
Kolmogorov	0.780	0.824	0.894	0.957	1.036
Pearson χ^2 with $k = 9$ and AOG	9.563	10.735	12.688	14.521	16.905
Pearson χ^2 with $k = 9$ and EPG	9.890	11.096	13.004	14.912	17.378
Pearson χ^2 with $k = 15$ and AOG	17.004	18.560	21.037	23.358	26.438
Pearson χ^2 with $k = 15$ and EPG	17.270	18.800	21.290	23.630	26.480

These conclusions are of the general nature, and this ordering is not rigid. Sometimes a test has an advantage in the power for some values of α and sample volumes n and loses for other values of α and n.

The attentive reader should notice that the power of nonparametric goodness-of-fit tests for testing complex hypotheses turns out considerably higher than the power of the same tests for testing simple hypotheses (with the same alternatives). This holds when using MLE. However, if we estimate the parameters as a result of minimizing the statistics of the tests themselves (find MD-estimates) then there is no growth in power in comparison with testing simple hypotheses [3].

We should note that the power of Pearson's χ^2 test while passing from testing simple hypotheses to testing complex hypotheses fails to increase. We should not forget that the power of χ^2-type tests (Pearson's and Nikulin's) depends not only on the hypotheses H_0 and H_1 and the volume sample n, but, given H_0 and H_1, also on the grouping method and the number of intervals. The number of intervals for which the power of the tests for a pair H_0 and H_1 is maximal depends on these hypotheses and the grouping method. The increase in the number of intervals need not always lead to increasing power of the χ^2-type tests [5].

For near hypotheses H_0 and H_1 while using the Pearson χ^2 test, the choice of AOG has a positive effect for both simple and complex hypotheses. However, this does not mean that the application of AOG always guarantees the maximal power of this test. For some not quite near hypotheses, another grouping method can turn out optimal, which can be found in result of maximizing the power of the test.

The conclusion on the unconditional positive effect of applying AOG cannot be extended to the Nikulin test: for the same pair of hypotheses H_0 and H_1 and one number k of intervals the test turns out more powerful with AOG, while for other k, with EPG. The dependence of the power on the grouping method turns out more complicated and requires further study.

In this article, we do not consider other modified χ^2-type tests, in particular, the Dzhaparidze–Nikulin test [10, 11], for which we can guess that it possesses about the same power as the Nikulin test. We should note that the interest in the modified χ^2-type tests has increased recently [12]. The reason is not only that the asymptotic distributions of the statistics of the tests are known, but also that the ideas underlying the construction of the statistics enable us to construct the corresponding tests for a wide range of problems where the application of other tests is impossible [13, 14]. Moreover, if necessary, we can always configure χ^2-type tests so that they distinguish a given pair of competing hypotheses H_0 and H_1 as well as possible by adjusting appropriately the subdivision into intervals and their number.

ACKNOWLEDGMENTS

The authors are grateful to D. M. Chibisov, Yu. N. Tyurin, and M.S. Nikulin for attention to the work and valuable comments which helped improve it.

The authors were supported by the Russian Foundation for Basic Research (project no. 06–01–00059–a).

REFERENCES

1. B. Yu. Lemeshko, S. B. Lemeshko, and S. N. Postovalov, "Comparative Analysis of the Power of Goodness-of-Fit Tests for Near Competing Hypotheses. I. The Verification of Simple Hypotheses," Sibirsk. Zh. Indust. Mat. **11** (2), 96–111 [J. Appl. Indust. Math. **3** (4), 462–475, 2009].
2. *Recommendation on Standardization. Applied Statistics. Rules of Check of Experimental and Theoretical Distribution of the Consent*, Vol. 2: *Nonparametric Goodness-of-Fit Test* (Izd. Standartov, Moscow, 2002) [in Russian].
3. B. Yu. Lemeshko and S. N. Postovalov, "Dependence of the Power and Distribution of Statistics for Nonparametric Tests on the Parameter Estimation Method," Zavodskaya Lab. Diagnostika Materialov **67** (7), 62–71 (2001).
4. G. V. Martynov, *Omega-Square Test* (Nauka, Moscow,1978) [in Russian].
5. B. Yu. Lemeshko and E. V. Chimitova, "On a Choice of the Number of Classes in χ^2 Goodness-of-Fit Tests," Zavodskaya Lab. Diagnostika Materialov **69** (1), 61–67 (2003).
6. B. Yu. Lemeshko and S. B. Lemeshko, "Comparative Analysis of the Tests for Deviations of a Probability Distribution from the Normal Distribution," Metrologiya, No. 2, 3–24 (2005).

7. V. G. Voinov, “On Optimality of the Rao—Robson—Nikulin Test,” Zavodskaya Lab. Diagnostika Materialov **72** (3), 65–70 (2006).
8. B. Yu. Lemeshko, S. N. Postovalov, and E. V. Chimitova, “On the Distribution and Power of the Nikulin χ^2 Test,” Zavodskaya Lab. Diagnostika Materialov **67** (3), 52–58 (2001).
9. B. Yu. Lemeshko and E. V. Chimitova, “Errors and Incorrect Procedures When Utilizing χ^2 Fitting Criteria,” Izmeritel’naya Tekhnika, No. 6, 5–11 (2002) [Measurement Techniques **45** (6), 572–581 (2002)].
10. K. O. Dzhaparidze and M. S. Nikulin, “On the Computation of Chi-Square-Type Statistics,” in *Problems of the Theory of Probabilistic Distributions*, Vol. 12 (Nauka, St. Petersburg, 1992), pp. 59–90.
11. K. O. Dzhaparidze and M. S. Nikulin, “On the Computation of Chi-Square-Type Statistics,” J. Math. Sci. **75** (5), 1910–1921 (1995).
12. V. Voinov, R. Alloyarova, and N. Pya, “A Modified Chi-Squared Goodness-of-Fit Test for the Three-Parameter Weibull Distribution and Its Applications in Reliability,” in *Mathematical Methods for Reliability, Survival Analysis and Quality of Life*, Ed. by C. Huber, M. Nikulin, N. Limnios, and M. Mesbah (Hermes, London, 2007), pp. 193–206.
13. V. Bagdonavicius and M. Nikulin, *Accelerated Life Models. Modeling and Statistical Analysis* (Chapman & Hall/CRC, Boca Raton, 2002).
14. L. Gerville-Reache, V. Couallier, and M. Nikulin, *Statistique des Essais Accélérés* (Hermes, London, 2007).